https://doi.org/10.46335/IJIES.2022.7.8.5 **e-ISSN: 2456-3463**
Vol. 7, No. 8, 2022, PP. 25-30

International Journal of Innovations in Engineering and Science, www.ijies.net

An Analytical Perspective on Various Deep Learning Techniques for Deep Fake Detection

Usha Kosarkar[1], Gopal Sakarkar[2], Shilpa Gedam[3]

[1]Assistant Professor, [2]Assistant Professor , [3]Assistant Professor
[1]Department of Science and Technology
G H Raisoni Institute of Engineering & Technology, Nagpur, India, 440028
[2]Department of Artificial Intelligence
G H Raisoni College of Engineering, Nagpur, India, 440016
[3]Department of Computer Science
Shivaji Science College, Nagpur, India, 440012

usha.kosarkar@raisoni.net

***Received on**: 11 June ,2022* ***Revised on**: 26 July ,2022,* ***Published on**: 01 August,2022*

Abstract *– The advent of deep fake technology has become a crucial concern in this digital world. A serious threat to an individual's privacy, democracy, and national security can be caused by deep fake. Deep fake algorithms can develop forgery multimedia content that we cannot distinguish from genuine ones. In this era of the cyber age, it has become seemingly difficult to identify between real digital content and fake content which are published across the Internet. It is a widely used technology used by cybercriminals to deceive security systems. If we are not cautious, deep fake technology can bring about a serious threat to the future of identity verification. There are many open-source and free software available to create deep fake content which makes it easy for amateurs to create technically brilliant digital content which is fake. On the other hand, many structured and efficient technologies have been developed to identify deep fakes. A few of the techniques available are like comparing the background, analyzing the pattern in the image, considering the blinking of the Eye, considering facial attributes, considering the head position, etc. This paper gives an introduction to deep fake, and a brief on deep fake creation and detection techniques..*

Keywords- Deep fake, Deep fake detection, GAN, CNN, Deep Learning, Security.

I- INTRODUCTION

In the ancient period, people spread foolish talk among the public and defame the people. In recent times automated video and image editing have a significant role in society. Spreading misinformation, fraud, and defamation are flowing nowadays. People are doing such various mischief to defame the people's character. Nowadays various tools are available to create falsified.

The word 'Deepfake' originated from 'Deep Learning', which belongs to machine learning methods and is based on AI [Artificial Intelligence][1]. Advances in Artificial Intelligence and the emergence of Generative Adversarial Networks [GAN] enabled the modeling and widespread of multiple techniques ableto bombard digital data, alter it or create its contents from scratch. This led to the birth of Deepfake technology. Using deepfake technology, one can easily swap the face of one person with another in any digital media[5].

The word 'Deepfake' originated from 'Deep Learning', which belongs to machine learning methods and is based on AI [Artificial Intelligence][1]. Advances in Artificial

https://doi.org/10.46335/IJIES.2022.7.8.5 **e-ISSN: 2456-3463**
Vol. 7, No. 8, 2022, PP. 25-30

International Journal of Innovations in Engineering and Science, www.ijies.net

Intelligence and the emergence of Generative Adversarial Networks [GAN] enabled the modeling and widespread of multiple techniques ableto bombard digital data, alter it or create its contents from scratch. This led to the birth of Deepfake technology. Using deepfake technology, one can easily swap the face of one person with another in any digital media[5].

The word 'Deepfake' originated from 'Deep Learning', which belongs to machine learning methods and is based on AI [Artificial Intelligence][1]. Advances in Artificial Intelligence and the emergence of Generative Adversarial Networks [GAN] enabled the modeling and widespread of multiple techniques ableto bombard digital data, alter it or create its contents from scratch. This led to the birth of Deepfake technology. Using deepfake technology, one can easily swap the face of one person with another in any digital media[5]. In recent years, automated video and picture editing has made significant advances. Deepfakes, or falsified videos with switched faces, have gotten a lot of interest because of their various potential users in fraud, defamation, entertainment, and spreading misinformation.

Examples of Deepfake

Deepfake technology can cause a grave impact on society because the manipulated media can spread rapidly in this digital world[8]. This technique can bring serious social issues when videos are manipulated to produce pornographic content of a celebrity or used badly in political arenas[4][12]. Deepfake videos involving election candidates can defame the candidate and cause a downfall in his political career. Even amateurs can deal with this technology without any profound computer knowledge. This will cross the social boundary and become a weapon of revenge. It is a very threatening tool when it comes into the hands of those who want to blackmail or defame others. In legal cases, evidence such as photos and videos are authentic sources and are highly critical in police investigations. Due to the manipulation caused by deepfake technology, these pieces of evidence are no longer trustworthy or reliable.[11]

II- DEEPFAKE CREATION

There are multiple techniques and tools for building deepfake content with the help of machine learning algorithms and deep learning. These algorithms are capable of generating content based on the data given as input. A deep-learning system can develop a compelling counterfeit by studying parameters from photos and videos of a target person from various angles, and then imitating its behavior Networks[1][2]

Going back to the emergence and usage of this technology, it was a Reddit user who first developed a popular Deepfake production, FakeApp. Autoencoder and decoder pairs were used for the content creation. In this method, the autoencoder extracted passive attributes of images based on face and the decoder rebuilt the face images. To switch the face in the image between source and target, two encoder-decoder pairs are required. Pair of encoder-decoder study the features of an image, and the attributes generated by the encoder are shared between the pair of encoder-decoder. This common encoder will study and try to get the similarity between the two image sets.

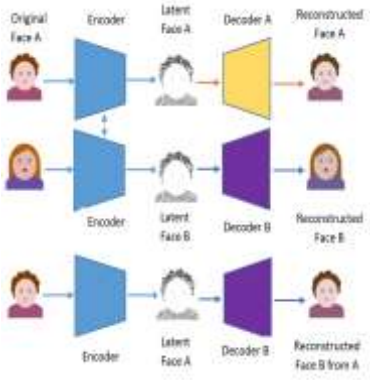
and speech patterns. Firstly, the program needs to be well informedand trained so that it can execute the tasks sequentially and create a new face or replace a part of a person's face. The program is provided a huge amount of data which is used for learning the context and then it creates its own, new data as required. They work based on autoencoders and GAN[Generative Adversarial

Step-wise Deepfake Creation Illustration

Step 1: Firstly, the image showing the original face needs to be chosen from the source video. This image is then fed as input to the Encoder or Deep Neural Network(DNN). Post-processing, a latent face image with passive attributes is created. This image is then given as an input to the Decoder. A reconstructed face image is developed for both the images(Face A and B).

Step 2: In this step, the DNN automatically develops an image that goes with the reconstructed face image.

Step 3: The DNN-created face image is now being placed in the original reference image and hence the Deepfake imageis created. Thereby the attribute generated from the face A image is supplied to the decoder B to rebuild face B attributes from the original face A image. The same concept is used in multiple Deepfake applications [1].



https://doi.org/10.46335/IJIES.2022.7.8.5 **e-ISSN: 2456-3463**
Vol. 7, No. 8, 2022, PP. 25-30

International Journal of Innovations in Engineering and Science, www.ijies.net

Fig 1- Illustration of Deepfake creation process using Two encoder-decoder pair

How to detect Deepfakes

Deepfake detection technologies are built on algorithms similar to deepfake creation algorithms. They check for indicators that wouldn't be present in real photos or videos (content).

Few signs of detection are:

1. Unusual blinking of eyes
2. Unusual skin color or skin tones
3. Speech a synchronization with lip movements.
4. Frame consisting of additional pixels
5. The lighting may not match with video
6. Audio mismatch
7. Consistent blur or flicker, low-resolution videos etc

II- DEEPFAKE DETECTION TECHNIQUES

Deepfakes are extremely dangerous and destructive if the privacy and security of society and democracy are considered. There are several methods for deepfake detection which are in prominence now. Previously deepfake detection methods used parameters by examining the various attributes and inconsistencies that are observed in the fake content. The current methods now are based on deep learning, and AI, and they can automatically generate primary and biased attributes to find deepfake content.

1. Background comparison

This is a very popular and easiest method for Deepfake detection. The first website to prohibit deepfake content was Gfycat, an online GIF repository platform. Gfycat mentioned that those videos were "objectionable". This platform removed deepfake videos by using facial acknowledgment models. It could identify the asymmetric attributes in the face of an uploaded video[After the primary identification, these clippings are then comparatively probed by hiding the facial attributes and then looking at the database in search of similar video with homogenous background and features[17]. Also, videos or content are assessed to check the facial similarity to conclude the reliability of the video. Coming the flaws of this model, it consumes a lot of time, resources, and a huge database to compare and validate the backgrounds. There are high chances that the background can be completely faked, and the blend of the new content(footage) would not be distinguished.

2. Analysis based on the temporal pattern

The behavior of any person can be distinguished by the collective order of their actions, emotions, gestures, etc. So sequence-based data can be considered as an approach to validate the video content. To analyze the temporal sequence of data, we can use a convolutional neural network (CNN) together with a Long Short-Term Memory (LSTM). In this method, each video frame is bypassed into a CNN and a series of highlight maps can be created for the LSTM[18]. Hence the network becomes capable to study the precise development-based practices of the content. Analyzing and studying human behavior in video motions is complicated since there are many activities, sub-activities, and frames involved in the content. The initial work is to identify the consecutive activities and limit them within temporal bounds. Tasks that occur at the same instance are captured.

Interconnected activity at a particular point in time may hamper the action of another individual due to cross-cutting.

This makes it difficult to learn the structure because the activity will have different deficient appearances which are difficult to identify. For the identification of video activity, different individual mixtures of CNNs are recommended. The prevailing models can examine the cropped videos it can analyze the short frames, and fragments, and record the motion details of each segment. These models are not capable to apply logical reasoning or do consecutive studies. This is meager when compared to human intelligence and capabilities. This highlights the fact that videos need not be considered fragments nor do they need to be handled independently. In the case of complex videos where there are multiple activities, the association among the activities needs to be considered and studied. Recurrent convolutional models are more significant regarding visual sequences. However, the performance on complex video datasets is quite less compared to that on single scene data. The reason behind this is that there is no thorough idea about reasoning by given temporal analysis and activity. Another flaw is that this model cannot figure out the extensive temporal structure. To study the interconnection between the actions in complicated videos, it is necessary to make use of another design. Also due to the decline and incline of noise, it is very tough to learn the characteristics of videos having long sequences. LSTM can be used to address this issue. It is a part of RNN.[15] Also, truncated backpropagation through time can solve this issue. Using this function, the gradient update happens from time to time. Using truncated backpropagation, a

https://doi.org/10.46335/IJIES.2022.7.8.5 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 25-30

sequence x is pulled apart into disjoint sub-sequences of length k. The "Interleaving" strategy that is introduced by Neuroscientists and psychologists contemplates related abilities or ideas in a parallel manner that trains the human mind. By making use of this interleaving concept, interleaved back diffusion based on time is proposed where interleaving connections are done which reduces the propagation length and results in increased learning of recurrent neural networks.

3. Consideration of Facial Attributes

Experts from UC Berkeley associated themselves with Adobe to develop a tool that can identify manually edited images by recognizing low-level facial distorting. Here, a CNN is produced using each case of images that were edited using the popular Face-Aware liquefy feature. The output of such approaches is welcoming, however since GAN-developed models are not available, this can only identify the manually edited images. So, this technology stands behind to fight against the modern-day Deepfakes [11]. The algorithm of Deepfake is capable to blend face images of a fixed size only. This is because these algorithms require more resources based on computation and processing time, sothey are capable of making lower resolutions of fixed sizes. As a result, these images would be subjected to raising and permutation, for illustration, resizing towards the postures of either the target facial coordinates that they would be succeeded in the synthesized composite In this procedure, the produced focus on various (ROI) and territory neighborhood are compared using a CNN model to determine which is the better choice. [10]The CNN model needs to be fed with data and trained for this purpose. The procedure needs to be well organized by replicating the resolution inconsistency in affine face warping squarely. It is necessary to first identify [13] the faces, characteristics, and indicators, after which the information must be extracted and inserted into the transform matrices to fit the standard design of both faces. To tune the face, Gaussian obscuring must be applied, and afterward whether the being is generated recovered from the asymmetrical distorted face by reversing the estimated modified grid estimation. A variety of scales are used to position the faces to assemble the information diversely, allowing for the acquisition or reproduction in ever-higher fidelity occurrences of the asymmetrical warped face.

4. Mesoscopic Analysis

Experts from the National Institute of Informatics (Tokyo) concluded that intense neural networks developed using an artificially low number of layers are suitable for identifying inconsistencies that are observed in Deepfakes. These deepfakes have high analytical powers and can bring about an accuracy of up to 90%. This method uses the Mesoscopic level of investigation to identify manually edited images of the face in videos. Microscopic analysis examinations depending on the sound of the image cannot be used ina compressed video. The human eye finds it difficult to identify furnished images, especially when the face of the image is presented. This strategy prefers to use the intermediate method by using a deep neural network with a smaller number of layers. In this module, the output of a few convolutional layers is put together with multiple kernel shapes. Hence the capacity space is increased and as an outcome, the model is also enhanced. It uses [3×3] enlarged convolutions in place of [5×5] original module convolutions to prevent high semantics [17]. By making use of enlarged convolutions,the initiation module can be fixed to control multi-scale data. That is, convolutions of [1x1] are used before expanded convolutions for measurement reduction, and an additional convolution of [1x1] is used in parallel to operate as a skip-association betweenprogressive modules.

5. Head Pose Estimation

In head pose estimation, the position of the human head is identified and validated. For the real-time head estimation, we require 2D, and 3D coordinates of facial landmarks and camera parameters [8]. Here, a support vector machine (SVM) classifier is used. An SVM classifier is evaluated and trained by providing a large number of Deepfakes and original images [17].

The following methods are further implemented:

1) Firstly, the Face detector detects and identifies 68 facial landmarks from each video frame or image. For this DLib programming tool is used.

2) Then, these identical 68 attributes are supplied to OpenFace2 to create a default t 3D facial model. The entire face is assessed separately and the head poses are considered from the focal face site.

3) The rotation matrices and interpretation vectors are obtained. The differences obtained are compacted in the form of a vector. Classification is done based on the mean and standard deviation.

6. Blinking of Eye

This method recognizes Deepfake videos by probing into the blinking of an eye for the current content.

Normally, a functioning human blink occurs every 2 to 10 seconds and lasts between 0.1 and 0.4 seconds on average, depending on the individual. Capturing a snapshot of someone blinking is approximately 7.5

https://doi.org/10.46335/IJIES.2022.7.8.5 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 25-30

percent more likely than taking a photo of someone not blinking when the video time is an average of 1/30 second. Because the majority of photographs available on the internet do not depict a person with their eyes closed, the absence of blinking eyelids is a telltale symptom of a Deepfake film. Here, a CNN (LRCN), which is a hybrid of RNN and CNN, is used to detect closed and open eye states from past temporal knowledge data using the earlier chronological knowledge data. The temporal connection between continuous frames is included because the duration between the opening and closing of eyelid blinking is a temporal activity, and LRCN can memorize the long-term dynamics to repair the artifacts.

IV- DEEP LEARNING FOR DEEPFAKE DETECTION

Similar to neural networks [13, 14], deep learning is a machine learning method that is built on the same notion as neural networks [13, 14]. The term "deep learning" refers to the employment of numerous hidden layers in a neural network in deep learning. Drawing inspiration from artificial networks, the deep learning architecture employs an unbounded number of hidden layers with bounded sizes to extract more information from raw input data than is possible with traditional learning architectures. The complexity of the training data [6] is used to calculate the number of hidden layers that should be used. For more complex data, more hidden layers are required to efficiently deliver the correct results. Since its inception, deep learning has been effectively applied in a range of fields, including computer vision, speech recognition, audio processing, automatic translation, and natural language processing. When deep learning is used in these disciplines, the results are state-of-the-art when compared to traditional machine learning methodologies. Deep learning has also demonstrated encouraging results in the detection of deepfakes. Many deep learning techniques have been proposed in the literature, including 1) convolutional neural network (CNN); 2) recurrent neural network (RNN); 3) long short-term memory (LSTM). The following are some examples: (LSTM). Following Table 1 shows the accuracy observed from the literature for the contents of the above-mentioned algorithms.

Table 1- Accuracy for Deep Learning Algorithms

Algorithm	Accuracy (%)
CNN	93.3
RNN	91.9
LSTM	89.3

Also, it has been creating new challenges and threats for cybersecurity authorities. Since deepfake technology is evolving and upgrading with time, it is a need to stay cautious and well-informed. There are several detection techniques and strategies available for deepfake detection.

As the deepfake algorithms are updating with time, it's necessary to introduce more powerful and efficient detection techniques as well. As a general public, we need to improve our ability to scrutinize, exercise, and evaluate our capabilities to judge the data which we come across in our day-to-day life. Following figure 2 shows the comparative analysis of the above algorithms. From the figure, it can be concluded that the deep learning algorithms can lead to a major accuracy boost as compared to machine learning algorithms. Also, coupled with the optimization in feature selection and adding more features will boost the accuracy more.

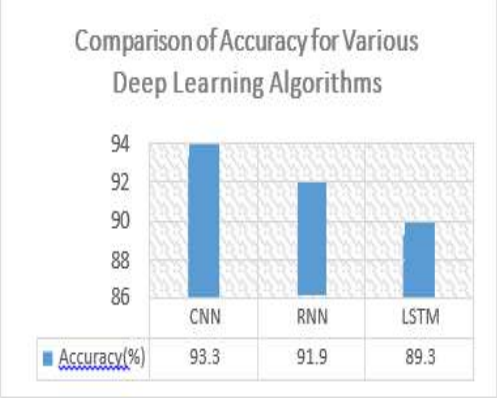


Fig 2.- Comparative Analysis of Various Deep Learning algorithms for DeepFake Detection

As a part of the study, a unique approach has been developed to reveal AI generated deepfake video together with powerful feature extraction and classification utilizing customized CNN. Comparing the existing model the proposed customized CNN outperforms two existing methods to get more testing accuracy.

Models	Accuracy
EfficientNetB7	86.98
EffiecientNetB1 + LSTM	86.02
ENSEMBLE	85.65
C-LSTM Xception	88.65
Xception_DFDC	87.45
DFDC_Rank90_CelebDF	80.32

Fig 3- Comparative Analysis of Various Models with

https://doi.org/10.46335/IJIES.2022.7.8.5 **e-ISSN: 2456-3463**
Vol. 7, No. 8, 2022, PP. 25-30

International Journal of Innovations in Engineering and Science, www.ijies.net

its accuracy

IV-CONCLUSION

Deepfakes have become highly critical these days as the technologies for deepfake creation are widely increasing so that even amateurs can get hold of it and create content within a short period. Social media is a large platform where these forgeries of digital content can spread quickly.

At the current time, it has become the most important and favorite tool of fraudsters and hackers for acquiring personal information from identity frauds. The algorithms used for deepfake creation are intelligent enough for independent decision-making. While this technology has few positive applications, especially in the field of entertainment and art, this has been causing serious threats in our society due to the widespread of fake digital.

REFERENCES

[1] *Thanh Thi Nguyen, Cuong M., Dung Tien Nguyen, Duc Thanh Nguyen, Saeid Nahavandi(2020), Deep Learning for Deepfakes Creation and Detection: A Survey, arXiv:1909.11573v2.*

[2] *Hrisha Y., Akshit K., Prakruti J. (2020). A Brief Study on Deepfakes, International Research Journal of Engineering and Technology (IRJET).*

[3] *Siwei L. (2021). Deepfake Detection: Current Challenges and Next Steps, 978-1-7281-1485-9/20/$31.00c 2020IEEE.*

[4] *Francesco Marra, Diego Gragnaniello, Davide Cozzolino, Luisa Verdoliva(2018). Detection of GAN-generated Fake Images over Social Networks, IEEE Conference on Multimedia Information Processing and Retrieval.*

[5] *Teng Zhang, Lirui Deng, Liang Zhang, Xianglei Dang(2020). Deep Learning in Face Synthesis: A Survey on Deepfakes, 2020 IEEE3rd International Conference on Computer andCommunication Engineering Technology.*

[6] *Deng Pan, Lixian Sun, Rui Wang, Xingjian Zhang, Richard O. Sinnott(2020). Deepfake Detection through Deep Learning, IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT).*

[7] *Chih-Chung Hsu , Yi-Xiu Zhuang and Chia-Yen Lee (2020). Deep Fake Image Detection Based on Pairwise Learning, AppliedScience.*

[8] *Nikita S. Ivanov, Anton V. Arzhskov, Vitaliy G. Ivanenko(2020). Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection, 978-1-7281-5761- 0/20/$31.00 ©2020 IEEE.*

[9] *Badhrinarayan Malolan, Ankit Parekh, Faruk Kazi(2020). Explainable Deep-Fake Detection Using Visual Interpretability Methods,3rd International Conference on Information and Computer Technologies(ICICT).*

[10] *Daniel Mas Montserrat, Hanxiang Hao, S. K. Yarlagadda, Sriram Baireddy, Ruiting Shao Janos Horv ´ ath, Emily Bartusiak, Justin Yang, David G ´ Uera, Fengqing Zhu, Edward*

[11] *J. Delp (2020). Deepfakes Detection with Automatic Face Weighting, Conference on Computer Vision and Pattern RecognitionWorkshops (CVPRW).*

[12] *Md. Shohel Rana, Andrew H. Sung(2020). DeepfakeStack: A Deep Ensemble-based Learning Technique for Deepfake Detection, International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom.*

[13] *Luca Guarnera, Oliver Giudice, and Sebastiano Battiato (2020). Fighting Deepfake by Exposing the Convolutional Traces on Images, IEEE Access DOI: 10.1109/ACCESS.2020.3023037*

[14] *Ali Khodabakhsh, Christoph Busch (2021). A Generalizable Deepfake Detector based on Neural Conditional Distribution Modelling, IEEE Xplore.*

[15] *Kui Zhu, Bin Wu (2020). Deepfake Detection with Clustering-based Embedding Regularization, IEEE Fifth International Conference on DataScience in Cyberspace.*

[16] *Dafeng Gong, Yogan Jaya Kumar, Ong Sing Goh, Zi Ye, Wanle Chi (2021), DeepfakeNet, an Efficient Deepfake Detection Method, (IJACSA) International Journal of Advanced Computer Science and Applications.*

[17] *Yang Wang(2020). A Mathematical Introduction to generate adversarial NETS (GAN), arXiv:2009.00169v1.*

[18] *Mohammed A. Younus, Taha M. Hasan(2020).Abbreviated view of Deepfake Videos Detection Techniques, International engineering conference" Sustainable Technology and Development".*

[19] *Bismi Fathima Nasar, Sajini T, Elizabeth Rose Lason(2020).Deepfake Detection in Media Files- Audios, Images and videos, IEEE Recent Advances in Intelligent Computational Systems(RAICS).*

[20] *Worku Muluye Wubet(2020), The Deepfake Challenges and DeepFake Video Detection, International Journal of Innovative Technology and Exploring Engineering(IJITEE)*

[21] *Tharindu Fernando, Clinton Fookes, Simon Denman, Sridha Sridharan(2021). Detection of Fake and Fradulent Faces via Neural Memory Networks, IEEE Transactions on Information Forensics and Security.*

[22] *Asad Malik, Minoru Kuribayashi,Sani M. Abdullahi, Ahmad Neyaz Khan(2022). DeepFake Detection for Human Face Images and Videos: A Survey,IEEE.*

[23] *Banu Priya M, Jhosiah Felips Daniel(2022). First Order Motion Model for Image Animation and Deep Fake Detection, Internation Conference on Computer Communication and Informatics(ICCCI).*